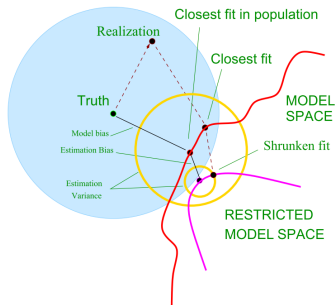
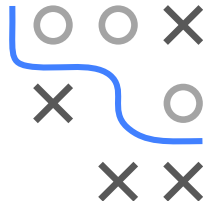


Bias-variance Tradeoff



- Understand the bias-variance trade-off
- Know the definition of model bias, estimation bias, and estimation variance



BIAS-VARIANCE TRADEOFF

In this slide set, we will visualize the bias-variance trade-off.

First, we start with a DGP \mathbb{P}_{xy} and a suitable loss function $L : \mathbb{R}^g \times \mathbb{R}^g \rightarrow \mathbb{R}$ where \mathbb{R}^g is numerical encoding of \mathcal{Y} . We measure the distance between models $f : \mathcal{X} \rightarrow \mathbb{R}^g$ via

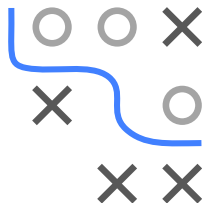
$$d(f, f') = \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_{\mathbf{x}}} [L(f(\mathbf{x}), f'(\mathbf{x}))].$$

We restrict our attention to losses for which d becomes a metric, e.g., L1-loss, L2-loss, etc.

We define f_{true} as the risk minimizer such that

$$f_{\text{true}} \in \arg \min_{f \in \mathcal{H}_0} \mathbb{E}_{(\mathbf{x}, y) \sim \mathbb{P}_{xy}} [L(y, f(\mathbf{x}))]$$

where $\mathcal{H}_0 = \{f : \mathcal{X} \rightarrow \mathbb{R}^g \mid d(\underline{0}, f) < \infty\}$ and $\underline{0} : \mathcal{X} \rightarrow \{0\}$.



BIAS-VARIANCE TRADEOFF

In practice, our model space \mathcal{H} usually is a proper subset of \mathcal{H}_0 and in general $f_{\text{true}} \notin \mathcal{H}$.

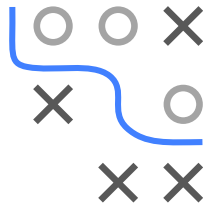
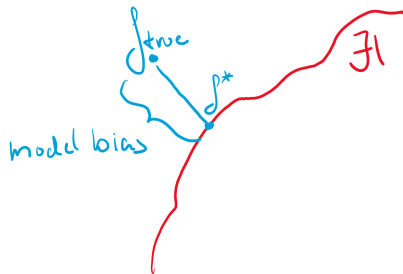
We define f^* as the risk minimizer in \mathcal{H} , i.e.,

$$f^* \in \arg \min_{f \in \mathcal{H}} \mathbb{E}_{(\mathbf{x}, y) \sim \mathbb{P}_{xy}} [L(f(\mathbf{x}, y))].$$

It is the function in \mathcal{H} closest to f_{true} , and we call $d(f_{\text{true}}, f^*)$ the model bias.



BIAS-VARIANCE TRADEOFF



BIAS-VARIANCE TRADEOFF

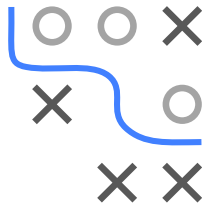
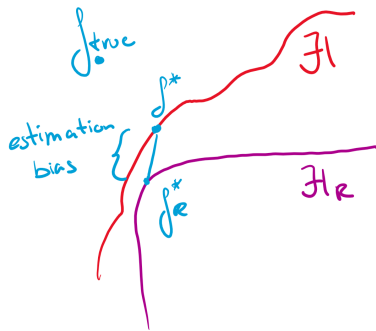
We can further restrict the model space such that \mathcal{H}_R is a proper subset of \mathcal{H} . We define f_R^* as the risk minimizer in \mathcal{H}_R , i.e.,

$$f_R^* \in \arg \min_{f \in \mathcal{H}_R} \mathbb{E}_{(\mathbf{x}, y) \sim \mathbb{P}_{xy}} [L(f(\mathbf{x}, y))].$$

It is the function in \mathcal{H}_R closest to f_{true} , and we call $d(f_R^*, f^*)$ the estimation bias.



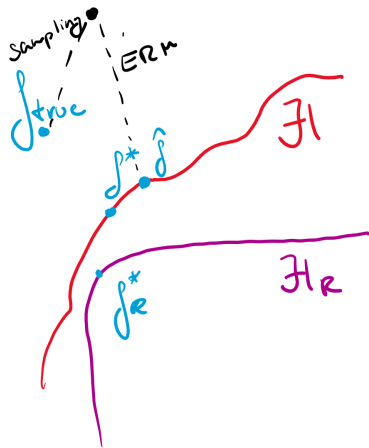
BIAS-VARIANCE TRADEOFF



BIAS-VARIANCE TRADEOFF

We sample a finite dataset $\mathcal{D} = (\mathbf{x}^{(i)}, y^{(i)})^n \in (\mathbb{P}_{\mathbf{xy}})^n$ and find via ERM

$$\hat{f} \in \arg \min_{f \in \mathcal{H}} \sum_{i=1}^n L(y^{(i)}, \hat{f}(\mathbf{x}^{(i)})).$$



Note:

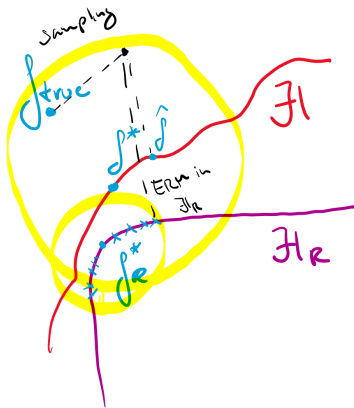
- $L : \mathcal{Y} \times \mathbb{R}^g \rightarrow \mathbb{R}$ is overloaded.
- The samples are only shown in the visualization for didactic purposes but are not an element of \mathcal{H} .



BIAS-VARIANCE TRADEOFF

We repeat the previous construction in the restricted model space \mathcal{H}_R and sample \hat{f}_R such that

$$\hat{f}_R \in \arg \min_{f \in \mathcal{H}_R} \sum_{i=1}^n L\left(y^{(i)}, \hat{f}(\mathbf{x}^{(i)})\right).$$



- We can measure the spread of sampled \hat{f}_R around f_R^* via $\delta = \text{Var}_{\mathcal{D}} \left[d(f^*, \hat{f}_R) \right]$ which we also call estimation variance.
- We observe that the increased bias results in a smaller estimation variance in \mathcal{H}_R compared to \mathcal{H} .

